Course Title: Big Data for Business Applications
Course Number: ISM6562
Credit Hours: 3

Instruction: Dr. Tim Smith

University Course Description:

The course will cover data management and data analytics for businesses using various big data technologies such as No-SQL databases, distributed file systems, MapReduce, and Spark.

Course Description:

With the advent of social media and IoT (Internet of Things), the data volume in organizations has increased rapidly in recent years. With the increase in data volume to terabytes and even petabytes, traditional database, and analytical techniques are insufficient. In this course, students will learn various big data technologies and how they can be used for data management and data analytics purposes to handle such massive datasets. The first half of the course will focus on big data storage technologies such as No-SQL databases and distributed systems. The second half of the course will focus on big data computational platforms such as Hadoop map-reduce and Spark. The course will cover in-depth spark programming on the big data platform.

Course Objectives:

- Review and reinforce SQL/RDBMS knowledge
- Understand basic terminal/shell commands
- Understand version control and collaboration using Git/GitHub.
- Identify the need for NoSQL/Non-RDBMS data systems
- Identify various components of big data technologies;
- Explain how big Data technologies can be used;
- Design and develop analytics applications on big data.
- Apply the knowledge of big data technologies to design and scale analytics applications

Learning Outcomes:

Upon completion of this course, students should be able to:

- Design and manage large business applications with millions of users
- Design analytic applications using several tens to hundreds of terabytes of data

- Develop business applications running on top of several tens to hundreds of terabytes of data

Perquisites:

- Completion of Data Mining
- Knowledge of SQL and Relational Databases
- Python Programming knowledge (Python DataCamp course will be offered as part of the class)
- Willingness to work hard with Big Data technology stack:
  - You will often need to conduct your own research and experiments to understand the technology and overcome obstacles more fully (this is an important skill to develop)

Instructor Contact Information and Communication

Email is generally the best way to contact me, and I will try to answer emails within 24 business hours. Please include the course number and section in the title of your email. Also, clarify your question/request – vagueness will result in wasting time with back-and-forth emails.

Instructor: Dr. Tim Smith
Office: School of Information Systems and Management (CIS Building)
Email: smith515@gmail.com

First-Day Attendance Policy

A student absent on the first day of the class will automatically be dropped. Attendance will be made during the first day of class, at the start of class.

Attending the class Virtually or Physically

Some class material is pre-recorded, and attendance is virtual. Other portions of the material will be delivered in class "face-to-face." For each week, you watch any assigned videos at your own pace, but it's important that you allocate time every day.

Course Materials:

Students do not need to purchase any software or book for this class. However, the instructor will direct students to various online materials, books, manuals, videos, and software to supplement the class lecture.

Book

1. Frank Kane's Taming Big Data with Apache Spark and Python - eBook freely available from the library

1. **Frank Kane's Taming big data with Apache Spark** and **Python** : real-world examples to help you analyze large datasets **with Apache Spark**

Frank Kane.

Author: Kane, **Frank**

Published: Birmingham, UK : Packt Publishing, 2017.

🅴 eBook

Ebook Central (EBL) USF ONLINE ACCESS Click to View

USF ONLINE RESOURCE General Collection

📁 add 🖨 print ✉ email 📄 txt 🌐➡ export

2. Ullman – Mining of Massive Datasets - http://www.mmds.org/

3. BOOK
PySpark Cookbook : Over 60 Recipes for Implementing Big Data Processing and Analytics Using Apache Spark and Python.
Drabas, Tomasz.; Lee, Denny.
Birmingham : Packt Publishing Ltd; 2018

BOOK
PySpark Cookbook : Over 60 Recipes for Implementing Big Data Processing and Analytics Using Apache Spark and Python.
Drabas, Tomasz.; Lee, Denny.
Birmingham : Packt Publishing Ltd; 2018

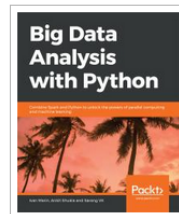📖 Available at USF ONLINE RESOURCE(SFPDA) General Collection 2 ›
🔗 Available Online ›

4. Big Data Analysis with Python: Combine Spark and Python to Unlock the Power of Parallel Computing and Machine Learning
by Ivan Marin, Ankit Shukla, and Sarang VK
Publisher
Packt Publishing, Limited

Big Data Analysis with Python : Combine Spark and Python to Unlock the Powers of Parallel Computing and Machine Learning

Ivan Marin, Ankit Shukla, and Sarang VK

Availability

Get all pages, require free third-party software, You have an active loan that expires in 0 day(s) and 23 hour(s).

Book Details

TITLE
Big Data Analysis with Python

How to Succeed in this course
- Access course regularly
- Watch all recorded videos
- Read all mandatory reading materials
- Practice the examples and assignments – this is very important!

- Combine knowledge from other courses such as Data Mining, Analytical Methods for Business, Statistical Programming, and Statistical Data Mining.

Grading Scale and Basis for Grades:

The grading will be based on the following components in the class –

Quizzes	20%
Assignments	24%
Data Camp Courses	6%
Final Project	20%
Mid-Term Exam	20%
Final Exam	30%

The students will be given an "Incomplete" grade only as per the university policy without any exception.

The final grading will be based roughly on the following scale –

>=97	A+
>=93	A
>=90	A-
>=87	B+
>=83	B
>=80	B-
>=77	C+
>=73	C
>=70	C-
>=67	D+
>=63	D
>=60	D-
Below 60%	F

Please note that this tentative grading scale can change based on the students and overall class performance. The faculty reserves the right to change the grading scale as he deems appropriate based on the class's overall performance.

Course Outline and Schedule:

The course is split into six weeks of work. Each week's material will be launched on Monday at 8:am, and the week's deliverables will be due Sunday evenings before 11:59 pm. The classes are asynchronous, each week will consist of watching videos, reading any assigned material, practicing, and writing quizzes and assignments.

Week	Discussion Topic
1 Aug 23	Introduction to Big Data Application Big data Storage Big data computation Complete DataCamp Courses: Command Shell & GitHub
2 Aug 30	Review Python Fundamentals Introduction to the Development Environment Tutorial: Creating a multi-processor Python Program Complete DataCamp Course: Intro to SQL SQL in Python
3 Sept 6	Online (no in-person class this week): • 2.1 What is big data • 2.2 What is no-SQL • 2.3 CAP Theorem • 2.4 Variants of NoSQL databases • 2.5 Cassandra Database
4 Sept 13	Hands-on with Cassandra
5 Sept 20	Hands-on with MongoDB Midterm Review
6 Sept 27	MidTerm Exam
7	HDFS, Hive, Impala MapReduce
8	Spark – Introduction and Fundamentals Spark – installing and working with PySpark in Jupyter Notebooks
9	Spark ML
10	Spark ML
11	Free Class – Work on Final Project
12	Final Exam Project Presentations – YouTube

Group Project: Big data analytics project

The final project will be a group project. You will be split into groups of 4-5 students. As a group, you are to identify a publicly available dataset and do the analysis, visualization, and model creation on the dataset using Jupyter Notebook and PySpark, and SparkML. More details will be provided during the semester. At the end of the group project, each team member will evaluate the performance of their other team members. This is called a **peer evaluation**. Your actual mark on the final project will be the team's mark multiplied by your percentage peer evaluation. For instance, let's say your team received 100%, but your peer evaluation was 9/10. This means that your mark for the final project will be 90%.

Standard University Policies

Policies about disability access, religious observances, academic grievances, academic integrity and misconduct, academic continuity, food insecurity, and sexual harassment are governed by a central set of policies that apply to all classes at USF. These may be accessed at:
https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx

COURSE POLICIES

Late Work Policy:

No credit will be given to late submissions. It's important that you plan your time well. You should target to complete the submission at least one day before the deadline. This is just good practice (and will help you in a professional setting in the future), as it will save you embarrassing situations such as – your computer wasn't working, your internet wasn't working, you had last-minute issues, etc. Such excuses will not play well in a professional setting and, therefore, will not be acceptable in this class.

Extra Credit Policy:

There are no opportunities for extra credit in this course. Student focus should be on the primary work in the course.

Grades of "Incomplete":

An "I" grade may be awarded to a student when 1) arrangements are made before the end of the semester, 2) in the judgment of the instructor, a valid reason is offered for granting an Incomplete, 3) a clear path to a standard grade is agreed to by the instructor and the student which will result in the successful completion of course requirements by the end of the succeeding semester. Offer specifics about your policy on incomplete grades. "I" grades not removed by the end of the next semester will be changed to "IF."

Email:

The primary means of communication between the instructor and students will be email. The instructor will occasionally send announcements to all students via Canvas. Students can feel free to email their instructor with questions at any time. Be sure to communicate clearly: What course, what section, and what is your question? Please anticipate a response time of 24 hours to email queries.

Canvas:

This course will use Canvas to disseminate materials, quizzes, tests, and midterm and final work. If you need help learning how to perform various tasks related to this course or other courses offered in Canvas, please consult the Canvas help guides. Contact USF's IT department at (813) 974-1222 or help@usf.edu if you have any technical issues.

Class Videos:

Audio and/or video recordings of lectures are allowable as per Florida law only for personal use. Such recordings should not include audio or video of other students in the class. Such recordings should not be publicly available and distributable as per Florida law.

Academic Integrity:

Academic integrity is the foundation of the University of South Florida System's commitment to the academic honesty and personal integrity of its university community. Academic integrity is grounded in certain fundamental values, which include honesty, respect, and fairness. Broadly defined, academic honesty is the completion of all academic endeavors and claims of scholarly knowledge as representative of one's own efforts. The final decision on an academic integrity violation and related academic sanction at any USF System institution shall affect and be applied to the student's academic status throughout the USF System unless otherwise determined by the independently accredited institution.

Disruption to Academic Process:

Disruptive students in the academic setting hinder the educational process. Disruption of the academic process is defined as the act, words, or general conduct of a student in a classroom or other academic environment which, in the reasonable estimation of the instructor: (a) directs attention away from the academic matters at hand, such as noisy distractions, persistent, disrespectful or abusive interruption of lecture, exam, academic discussion, or general University operations, or (b) presents a danger to the health, safety, or well-being of self or other persons.

Student Academic Grievance Procedures:

These procedures provide all undergraduate and graduate students taking courses within the University of South Florida System an opportunity to objectively review facts and events pertinent to the cause of

the academic grievance. An "academic grievance" is a claim that a specific academic decision or action that affects that student's academic record or status has violated published policies and procedures or been applied to the grievant differently from that used for other students.

Disability Access:

Policies about disability access, religious observances, academic grievances, academic integrity and misconduct, academic continuity, food insecurity, and sexual harassment are governed by a central set of policies that apply to all classes at USF. These may be accessed at: https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx

Sexual Misconduct/Sexual Harassment Reporting:

USF is committed to providing an environment free from sex discrimination, including sexual harassment and sexual violence (USF System Policy 0-004). The USF Center for Victim Advocacy and Violence Prevention is a confidential resource where you can talk about incidents of sexual harassment and gender-based crimes, including sexual assault, stalking, and domestic/relationship violence. This confidential resource can help you without having to report your situation to either the Office of Student Rights and Responsibilities (OSSR) or the Office of Diversity, Inclusion, and Equal Opportunity (DIEO), unless you request that they make a report. Please be aware that in compliance with Title IX and under the USF System Policy, educators must report incidents of sexual harassment and gender-based crimes, including sexual assault, stalking, and domestic/relationship violence. If you disclose any of these situations in class, in papers, or to me personally, I am required to report it to OSSR or DIEO for investigation. Contact the USF Center for Victim Advocacy and Violence Prevention: (813) 974-5757.

Religious Observances:

All students have a right to expect that the University will reasonably accommodate their religious observances, practices, and beliefs. If you observe religious holidays, you should plan your allowed absences to include those dates.

Incomplete Grade Policy:

An Incomplete grade ("I") is exceptional and granted at the instructor's discretion only when students are unable to complete course requirements due to illness or other circumstances beyond their control. The course instructor and student must complete and sign the "I" Grade Contract Form that describes the work to be completed, the date it is due, and the grade the student would earn factoring in a zero for all incomplete assignments. The due date can be negotiated and extended by student/instructor as long as it does not exceed two semesters for undergraduate courses and one semester for graduate courses from the original date grades were due for that course. An "I" grade not cleared within the two

semesters for undergraduate courses and one semester for graduate courses (including summer semester) will revert to the grade noted on the contract.

Covid-19 Procedures

All students must comply with university policies and posted signs regarding COVID-19 mitigation measures (https://www.usf.edu/coronavirus/).

Additional details are available on the University's Core Syllabus Policy Statements page: https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx

Online Proctoring

Quizzes and exams will be conducted using online proctoring tools. Keeping the audio and video (microphone and camera) on during such exams and quizzes is necessary. If the student is unwilling to use these, the student is asked not to register for this course. Any student may elect to drop or withdraw from this course before the end of the drop/add period. Online exams and quizzes within this course may require online proctoring. Therefore, students must have a webcam (USB or internal) with a microphone when taking an exam or quiz. Students understand that this remote recording device is purchased and controlled by the student and that recordings from any private residence must be done with any person's permission. To avoid concerns, students should select private spaces for testing or in certain cases, the classroom. The University library and other academic sites at the University offer secure private settings for recordings, and students with concerns may discuss the location of an appropriate space for the recordings with their instructor or advisor. Students must ensure that recordings do not invade any third-party privacy rights and accept all responsibility and liability for violations of third-party privacy concerns. Setup information will be provided before taking the proctored exam. For additional information about online proctoring, you can visit the online proctoring student FAQ at https://www.usf.edu/innovative-education/student-resources/index.aspx

Turnitin.com

In this course, turnitin.com may be utilized. Turnitin is an automated system that instructors may use to quickly and easily compare each student's assignment with billions of websites and an enormous database of student papers that grows with each submission. Accordingly, you will be expected to submit all assignments in both hard copy and electronic format. After the assignment is processed, as an instructor, I receive a report from turnitin.com that states if and how another author's work was used in the assignment. For a more detailed look at this process, visit http://www.turnitin.com. Essays are due at turnitin.com the same day as in class.

Netiquette Guidelines:

1. Act professionally in the way you communicate. Treat your instructors and peers respectfully like you would in a face-to-face environment. Respect other people's ideas and be constructive when explaining your views about points you may disagree with.
2. Be sensitive. Be respectful and sensitive when sharing your ideas and opinions. People in your class will have different linguistic backgrounds, political and religious beliefs, or other general differences.
3. Proofread and check spelling. Doing this before sending an email or posting a thread on a discussion board will allow you to make sure your message is clear and thoughtful. Avoid using all capital letters, it can be perceived as if you are shouting, and it is more difficult to read.
4. Keep your communications focused and stay on topic. Complete your ideas before changing the subject. By keeping the message on focus, you allow the readers to easily get your idea or answers they are looking for.
5. Be clear with your message. Avoid using humor or sarcasm. Since people can't see your expressions or hear your tone of voice, meaning can be misinterpreted.

Email and Discussion Board Guidelines:

1. Use the subject line effectively using a meaningful line in your email or discussion.
2. Keep your emails and postings related to the course content. Unless the instructor requests, you should not post anything personal on a discussion board.
3. Any personal, course or confidential issues should be emailed to the instructor. The discussion boards are public spaces; therefore, personal issues should not be posted there.
4. Posts and discussions unrelated to class content are prohibited. A warning will be given for such a posting. Repeat offenders will lose access to the online class materials and thus risk failing the class.

End of Semester Student Evaluations:

All classes at USF use an online system for students to provide feedback to the University regarding the course. These surveys will be made available at the end of the semester, and the University will notify you by email when the response window opens. Your participation is highly encouraged and valued.

WhatsApp, GroupMe, and Student-to-Student Communication:

While students may use digital communication tools (WhatsApp, GroupMe, etc.) to communicate with fellow students, it is important to remember that academic integrity policies still apply in these environments. Informing others about the contents of tests is prohibited by the official regulation, as is receiving unauthorized information about an examination. Students are expected and required to report instances of such violations to the instructor immediately.

Laptop:

You should have your own laptop with at least 8 GB RAM to do the work needed in the class.

Title IX Policy:

Title IX provides federal protections for discrimination based on sex, which includes discrimination based on pregnancy, sexual harassment, and interpersonal violence. In an effort to provide support and equal access, **USF has designated all faculty (TA, Adjunct, etc.) as Responsible Employees who are required to report any disclosures of sexual harassment, sexual violence, relationship violence, or stalking.** The Title IX Office makes every effort, when safe to do so, to reach out and provide resources and accommodation and to discuss possible options for resolution. Anyone wishing to make a Title IX report or seeking accommodations may do so online, in person, via phone, or email to the Title IX Office. For information about Title IX or for a full list of resources, please visit: https://www.usf.edu/title-ix/gethelp/resources.aspx. *If you are unsure what to do, please contact Victim Advocacy – a confidential resource that can review all your options – at 813-974-5756 or* va@admin.usf.edu*.*

Course Hero / Chegg Policy:

The USF Policy on Academic Integrity specifies that students may not use websites that enable cheating, such as by uploading or downloading material for this purpose. This does apply specifically to Chegg.com and CourseHero.com – almost any use of these websites (including uploading proprietary materials) constitutes a violation of the academic integrity policy.

Academic Accommodations

Students with disabilities are responsible for registering with Student Accessibility Services (SAS) to receive academic accommodations. You can visit the SAS website for additional information about academic accommodations and resources.
Academic Support Services

The USF Office of Student Success coordinates and promotes university-wide efforts to enhance undergraduate and graduate student success. Please visit the Office of Student Success website for a comprehensive list of academic support services available to all USF students.

Canvas Technical Support

If you have technical difficulties in Canvas, you can find access to the Canvas guides and video resources in the "Canvas Help" page on the homepage of your Canvas course. You can also contact the help desk by calling 813-974-1222 in Tampa or emailing help@usf.edu.
IT website for the Tampa campus.
IT website for the St. Pete campus.
IT website for the Sarasota-Manatee campus.

Center for Victim Advocacy

ISM6562

The [Center for Victim Advocacy](#) empowers survivors of crime, violence, or abuse by promoting the restoration of decision making, by advocating for their rights, and by offering support and resources. Contact information is available online.

Counseling Center

The Counseling Center promotes the wellbeing of the campus community by providing culturally sensitive counseling, consultation, prevention, and training that enhances student academic and personal success. Contact information is available online.

> [Counseling Center website for the Tampa campus.](#)
> [Counseling Center website for the St. Pete campus.](#)
> [Counseling Center website for the Sarasota-Manatee campus.](#)

Campus Free Expression:

It is fundamental to the University of South Florida's mission to support an environment where divergent ideas, theories, and philosophies can be openly exchanged and critically evaluated. Consistent with these principles, this course may involve discussion of ideas that you find uncomfortable, disagreeable, or even offensive.

In the instructional setting, ideas are intended to be presented in an objective manner and not as an endorsement of what you should personally believe. Objective means that the idea(s) presented can be tested by critical peer review and rigorous debate, and that the idea(s) is supported by credible research.

Not all ideas can be supported by objective methods or criteria. Regardless, you may decide that certain ideas are worthy of your personal belief. In this course, however, you may be asked to engage with complex ideas and to demonstrate an understanding of the ideas. Understanding an idea does not mean that you are required to believe it or agree with it.

USF Core Syllabus Policies

USF has a set of central policies related to student recording class sessions, academic integrity and grievances, student accessibility services, academic disruption, religious observances, academic continuity, food insecurity, and sexual harassment that **apply to all courses at USF**. Be sure to review these online at: https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx Any USF policy will supersede any policy mentioned in this syllabus.